

CLAIMS

What is claimed is:

1 1. A method for generating a checkpoint for a virtual machine (VM) in a
2 virtual computer system, the VM using a parent disk file and a set of VM memory, the
3 method comprising:

4 creating a copy-on-write (COW) disk file pointing to the parent disk file in
5 use by the VM;

6 stopping the VM, and while the VM is stopped:

7 marking the memory of the VM copy-on-write, the VM memory
8 constituting original VM memory;

9 saving substantially all of the device state of the VM to memory;

10 and

11 switching the VM to use the COW disk file instead of the parent
12 disk file;

13 resuming operation of the VM;

14 handling disk COW faults to the COW disk file;

15 handling memory COW faults to the original VM memory to generate
16 copies of the original VM memory for read and write use by the VM;

17 saving the device state from memory to a checkpoint data store; and

18 saving the original VM memory to the checkpoint data store.

1 2. The method of claim 1, wherein the VM is still running when the COW disk
2 file is created.

1 3. The method of claim 1, further comprising copying the parent disk file after
2 any pending disk writes complete, and using the copy of the parent disk file for the
3 checkpoint.

1 4. The method of claim 3, further comprising committing the COW disk file
2 into the original parent disk file.

1 5. The method of claim 3, wherein the step of committing the COW disk file
2 into the original parent disk file comprises creating one or more new COW disk files for
3 use by the VM while the COW disk file previously used by the VM is being committed.

1 6. The method of claim 3, wherein the copy of the parent disk file is
2 indicated for use for the checkpoint by adding a disk file pointer to the checkpoint file.

1 7. The method of claim 1, wherein the steps of creating the COW disk file
2 and handling disk COW faults are performed by a data storage device that is external to
3 the virtual computer system.

1 8. The method of claim 1, wherein the step of saving the original VM memory
2 to the checkpoint data store is delayed until all pending disk reads complete.

1 9. The method of claim 1, wherein the checkpoint data store comprises raw
2 data stored in a data storage medium.

1 10. The method of claim 1, wherein the checkpoint data store comprises a file
2 stored in a data storage medium.

1 11. The method of claim 10, wherein the data storage medium comprises a
2 disk drive.

1 12. The method of claim 10, wherein the data storage medium comprises
2 memory.

1 13. The method of claim 12, wherein the memory comprises high-speed
2 random access memory.

1 14. The method of claim 12, wherein the memory comprises flash memory.

1 15. The method of claim 1, further comprising, after the operation of the VM is
2 resumed, forcing memory COW faults on original VM memory that is affected by any
3 new disk reads, prior to issuing the new disk reads.

1 16. The method of claim 1, wherein, if there is a pending read followed by a
2 pending write to the same disk block, the resumption of the operation of the VM is
3 delayed until all pending disk operations complete.

1 17. The method of claim 16, further comprising, if the resumption of the
2 operation of the VM is not delayed until all pending disk operations complete, reissuing
3 any disk reads that affect any VM memory for which a COW fault has occurred.

1 18. The method of claim 1, wherein, for the step of handling disk COW faults
2 to the COW disk file, if there is a pending disk write to the same COW block group for
3 which the disk COW fault has occurred, the handling of the disk COW fault and the write
4 that caused the disk COW fault are delayed until the pending disk write completes.

1 19. A method for generating a checkpoint for a virtual machine (VM) in a
2 virtual computer system, the VM using a virtual disk and a set of VM memory, the
3 method comprising:

4 maintaining, in an unmodified state, the contents of the virtual disk at the
5 time for which the checkpoint is generated, while allowing the VM to continue
6 using the virtual disk;

7 saving substantially all of the device state of the VM, at the time for which
8 the checkpoint is generated, to a checkpoint data store; and

9 saving the set of VM memory, at the time for which the checkpoint is
10 generated, to the checkpoint data store by performing the following steps:

11 marking the set of VM memory as copy-on-write (COW), the set of
12 VM memory constituting original VM memory;

13 allowing the VM to continue using the VM memory;

14 responding to memory COW faults related to the VM memory by
15 generating copies of the original VM memory for read and write use by the
16 VM; and
17 saving the original VM memory to the checkpoint data store.

1 20. The method of claim 19, wherein the execution of the VM is stopped while
2 the VM memory is marked COW.

1 21. The method of claim 19, wherein the virtual disk is initially mapped to a
2 parent disk file on a physical disk and wherein the step of maintaining, in an unmodified
3 state, the contents of the virtual disk comprises creating a copy-on-write (COW) disk file
4 pointing to the parent disk file and mapping the virtual disk to the COW disk file.

1 22. The method of claim 19, wherein the step of saving the device state of the
2 VM to the checkpoint data store comprises saving the device state to memory while the
3 VM is not executing and copying the device state from memory to the checkpoint data
4 store after the VM has resumed execution.

1 23. A method for generating a checkpoint for a virtual machine (VM) in a
2 virtual computer system, the VM using a virtual disk and a set of VM memory, the
3 method comprising:
4 maintaining, in an unmodified state, the contents of the virtual disk at the
5 time for which the checkpoint is generated, while allowing the VM to continue
6 using the virtual disk, the unmodified contents of the virtual disk constituting a
7 checkpointed virtual disk, and the contents of the virtual disk used by the VM
8 constituting an ongoing virtual disk;
9 saving substantially all of the device state of the VM, at the time for which
10 the checkpoint is generated, to a checkpoint data store;
11 saving the contents of the VM memory, at the time for which the
12 checkpoint is generated, to the checkpoint data store, and allowing the VM to
13 continue using the VM memory, the contents of the VM memory saved to the

14 checkpoint data store constituting a checkpointed VM memory, and the contents
15 of the VM memory as used by the VM constituting an ongoing VM memory; and
16 allowing the VM to execute during at least a part of the time during which
17 the checkpoint is being generated, and ensuring that the results of any pending
18 disk writes are applied to both the checkpointed virtual disk and the ongoing
19 virtual disk, that the results of any new disk writes are applied to the ongoing
20 virtual disk, but not to the checkpointed virtual disk, that the results of any
21 pending disk reads are applied to both the checkpointed VM memory and the
22 ongoing VM memory, and that the results of any new disk reads are applied to
23 the ongoing VM memory, but not to the checkpointed VM memory.

1 24. The method of claim 23, wherein the step of saving the contents of the VM
2 memory to the checkpoint data store comprises:
3 marking the set of VM memory as copy-on-write (COW), the set of VM
4 memory constituting original VM memory;
5 allowing the VM to continue using the VM memory;
6 responding to memory COW faults related to the VM memory by
7 generating copies of the original VM memory for read and write use by the VM;
8 and
9 saving the original VM memory to the checkpoint data store.

1 25. The method of claim 24, wherein the step of ensuring that the results of
2 any pending disk reads are applied to the ongoing VM memory comprises reissuing any
3 pending disk reads for which the results of the read were directed to original VM
4 memory for which a COW fault has occurred, but directing the reissued disk reads to
5 the corresponding copies of the original VM memory instead of the original VM memory.

1 26. The method of claim 24, wherein the step of ensuring that the results of
2 any new disk reads are not applied to the checkpointed VM memory comprises forcing
3 COW faults for any original VM memory that would otherwise be affected by the new
4 disk reads.

1 27. The method of claim 23, wherein the virtual disk is initially mapped to a
2 parent disk file on a physical disk and wherein the step of maintaining, in an unmodified
3 state, the contents of the virtual disk comprises creating a copy-on-write (COW) disk file
4 pointing to the parent disk file and mapping the virtual disk to the COW disk file.

1 28. The method of claim 27, wherein the step of ensuring that the results of
2 any pending disk writes are applied to the ongoing virtual disk comprises, if there is a
3 pending disk write to the same COW block group as a subsequent write for which a disk
4 COW fault has occurred, delaying responding to the disk COW fault and delaying the
5 subsequent write until the pending disk write completes.

1